Speaker Recognition Menggunakan MFCC dan Algoritma DTW

Novarika Florencia*[1], Maulidina Tatiana[2], Rachmansyah[3] , Derry Alamsyah[4]
[1,2]**STMIK GI MDP**; **Jl. Rajawali No.14,+62(711)376400/376360**
[3]**Program Studi Teknik Informatika, STMIK GI MDP Palembang**
e-mail: ***[1]novclow@mhs.mdp.ac.id**, [2]**maulidinatania@mhs.mdp.ac.id**,
[3]**rachmansyah@mdp.ac.id**, [4]**derry@mdp.ac.id**

Abstrak

Speaker recognition (pengenalan penutur) adalah kemampuan sebuah mesin atau program untuk mengenali atau memastikan identitas penutur berdasarkan ciri suaranya. Ada dua tipe teks pada speaker recognition, yaitu text dependent dan text independent. Beberapa penelitian memverifikasi suara menggunakan Dynamic Time Warping (DTW) telah mendapat hasil yang baik. Begitu pula penelitian identifikasi suara dengan Mel Frequency Cepstral Coefficients (MFCC) dan algoritma GMM. Penelitian tersebut sebagaian besar menggunakan Bahasa Inggris, India, Persia, Tamil, dan Indonesia. Data ucapan umumnya mengambil sampel suara menuturkan sebuah kata atau beberapa kata dan bersifat teks dependent. Maka dari itu akan dilakukan penelitian pengenalan suara menggunakan metode MFCC dan algoritma DTW dengan tipe teks independent dan menuturkan beberapa kalimat berdurasi 30 detik. Data didapatkan dari hasil rekaman sendiri 20 orang penutur (10 laki-laki, 10 perempuan) pada tiga kondisi ruangan yang berbeda. Pengujian diketahui dengan menggunakan metode Maximum Likelihood Classification. Hasil pengujian menunjukkan kemampuan sistem memverifikasi penutur sebesar 70% untuk dimensi satu dan dua serta kemampuan mengidentifikasi penutur sebesar 15% untuk dimensi satu dan dua.

Kata kunci: *Speaker Recognition, Identification,* MLC, MFCC, DTW, Polinomial Kernel.

Abstract

Speaker recognition is the ability of a machine or program to identify or confirm the identity of the speaker based on voice characteristics. There are two types of text in speaker recognition, which is text-dependent and text-independent. Some studies verify sound using Dynamic Time Warping (DTW) has got a good result. Similarly, voice identification study with Mel Frequency cepstral Coefficients (MFCC) and GMM algorithm. The study uses a large part of English, Indian, Persian, Tamil, and Indonesia. Data speech generally take samples of voice utters a word or several words and text are dependent. Therefore the research will be done using voice recognition MFCC and DTW algorithm with independent text types and said a few sentences duration of 30 seconds. Data obtained from the recording itself 20 speakers (10 male, 10 female) at three different room conditions. Tests determined by using the method of Maximum Likelihood Classification. The test results demonstrate the ability of the system to verify the speaker by 70% for one and two dimension and the ability to identify speakers by 15% for one and two dimensions.

Keywords: *Speaker Recognition, Identification, MLC, MFCC, DTW, Polynomial Kernel.*

1. PENDAHULUAN

Speaker recognition adalah teknik yang mampu mengidentifikasi atau membuktikan identitas dari pembicara. Prosedur pengenalan berdasarkan parameter sampel ucapan dan dibagi menjadi dua bagian, *speaker verification* dan *speaker identification. Speaker identification* adalah proses mendapatkan identitas dari seorang penutur dengan membandingkan fitur-fitur suara yang diinputkan dengan semua fitur-fitur dari setiap penutur yang ada dalam *database. Speaker verification* adalah proses verifikasi seorang pembicara, di mana sebelumnya telah diketahui identitas penutur tersebut berdasarkan data yang telah diinputkan [1].

Speaker recognition juga diklasifikasikan ke dalam *text dependent* dan *text independent* [2]. Pada *text dependent*, penutur harus menuturkan ucapan yang sama saat *training* dan *testing*. Pada *text independent* ucapan saat proses *testing* berbeda dari ucapan saat *training* [3]. *Text dependent* biasa digunakan dalam absensi dan sensor kamera menggunakan suara, sedangkan *text independent* dapat digunakan untuk keamanan, misalnya kunci pengaman rumah atau brankas yang hanya bisa dibuka dengan suara sang pemilik. Hal ini dikarenakan setiap manusia memiliki sesuatu yang unik atau khas yang hanya dimiliki oleh dirinya sendiri, tidak mudah hilang, tidak mudah lupa dan melekat sehingga menjadi identitas dirinya [4].

Dua modul utama dalam *speaker recognition* adalah ekstaksi fitur (*feature extraction*) dan pencocokkan fitur (*pattern matching*). Proses ekstraksi mengambil sejumlah kecil sinyal suara yang kemudian digunakan untuk mewakili masing-masing penutur. Proses pencocokan melibatkan prosedur yang tepat untuk mengidentifikasi penutur yang tidak diketahui dengan membandingkan fitur yang diambil dari dataset penutur yang sudah diketahui [5].

Berdasarkan penelitian-penelitian yang telah dilakukan sebelumnya, fitur ekstraksi yang digunakan untuk sinyal suara antara lain *Linear Predictive Cepstral Coefficient* (LPCC) [2], *Mel Frequency Cepstral Coefficients* (MFCC) [1] dan lainnya. Dari beberapa penelitian dapat dilihat bahwa MFCC dinyatakan sebagai yang memiliki parameter karakteristik penting untuk mengenal suara dan menghasilkan tingkat akurasi yang baik. Lalu untuk fitur pencocokkan, algoritma *Dynamic Time Warping* (DTW) [6], *Gaussian Mixture Model* GMM [3] dan *Vector Quantization* (VQ) [2] adalah *template matching* yang paling sering digunakan, terutama untuk mengenali sinyal suara dalam kondisi rekaman di lingkungan berisik. Dari ketiganya penggunaan DTW mendapatkan hasil akurasi yang baik.

Maka dari itu dilakukan penelitian *text independent* dalam Bahasa Indonesia dengan menerapkan MFCC sebagai *feature extraction* dan algoritma DTW sebagai *template matching*. Penelitian dilakukan dalam tiga kondisi ruangan perpustakaan (sepi, *noiseless*, dan *noise*+musik) dengan melakukan proses *training* terlebih dahulu sehingga didapatkan tingkat akurasi yang baik.

2. METODE PENELITIAN

2.1 Studi Literatur

2.1.1 Penelitian Terkait

Pada penelitian [1] nilai akurasi yang didapat sebesar 93,254% untuk *speaker verification* dalam Bahasa Indonesia menggunakan *Mel Frequency Cepstral Coefficients* (MFCC) dan *Dynamic Time Warping* (DTW). Pada penelitian [6] menggunakan MFCC sebagai *feature extraction* dan algoritma *Algebraic Approach* (AA) dan DTW sebagai *template matching* untuk *speaker verification* dalam Bahasa Inggris. Hasil akurasi yang didapatkan sebesar 85,455%. Pada penelitian [2] dibuat program untuk *speaker identification* dalam Bahasa Persia dengan mengkombinasikan *feature extraction Linear Predictive Cepstral Coefficient* (LPCC), *Cepstral Mean Subtraction* (CMS) dan MFCC masing-masing dengan algoritma DTW dan *Vector Quantization* (VQ) dengan hasil akurasi antara 80-100%. Pada penelitian yang dilakukan [3] menggunakan

Gaussian Mixture Model (GMM) serta kombinasi MFCC dan *Dynamic* MFCC (DMFCC) sebagai *feature extraction* untuk *speaker identification* dalam Bahasa Tamil dan Inggris. Hasil pengujian mendapatkan nilai akurasi sebesar 98,8%.

2.1.2 *Mel Frequency Cepstral Coefficients* (MFCC)

MFCC merupakan salah satu fitur populer teknik ektrasi untuk sinyal *speech* (ucapan). Tujuan utama dari MFCC adalah untuk meniru tingkah laku pendengaran manusia yang tidak dapat menerima frekuensi diatas 1 Khz. MFCC didasarkan pada variasi *bandwidth* telinga manusia yang kritis dengan frekuensi. MFCC memiliki dua jenis filter yang bekerja secara linier pada frekuensi rendah di bawah 1000 Hz dan logaritmik diatas 1000 Hz. Sebuah nada subjektif diberikan pada Frekuensi Skala Mel untuk menangkap karakteristik penting dari fonetik dalam ucapan [7]. Beberapa keunggulan dari MFCC adalah [1]:

a. Mampu untuk menangkap karakteristik suara yang sangat penting bagi pengenalan suara, atau dengan kata lain dapat mengangkap informasi-informasi penting yang terkandung dalam *signal* suara.
b. Menghasilkan data seminimal mungkin, tanpa menghilangkan informasi-informasi penting yang dikandungnya.
c. Mereplikasi organ pendengaran manusia dalam melakukan persepsi terhadap *signal* suara.

Proses MFCC adalah sebagai berikut:

1. *Pre-emphasis*

Gelombang *digital* ucapan memiliki dinamik yang tinggi dan memilki kebisingan aditif. Untuk mengurangi dinamiknya dan secara spectral meratakan sinyal ucapan, dilakukanlah *pre-emphasis*. Bentuk yang digunakan dalam *pre-emphasis* adalah sebagai berikut [7]:

$$Y[n] = X[n] - aX[n-1] \quad (1)$$

Keterangan:
$Y[n]$ = signal hasil *pre-emphasis*
$X[n]$ = signal sebelum *pre-emphasis*
a = konstanta $0.9 \leq a \leq 1.0$

2. Framing

Proses ini adalah di mana sinyal suara dibagi menjadi *frame* N sampel. *Frame* yang berdekatan dipisahkan oleh M (M<N). Nilai-nilai tipikal yang digunakan adalah M=100 dan N=256. Proses ini dilakukan sampai seluruh sinyal dapat diproses [7].

3. *Windowing*

Pada tahap ini, dilakukan pemrosesan *window* pada masing-masing *frame* individual untuk meminimalisasi sinyal tak kontinu pada awal dan akhir masing-masing *frame*. *Window* didefinisikan sebagai W_n , $0 \leq n \leq N-1$ di mana N adalah jumlah sampel di setiap *frame*. Proses *windowing* dihitung dengan [7]:

$$Y_n = X_n \times W_n \quad (2)$$

Keterangan :

Y_n = sinyal hasil *windowing* sampel ke -n
X_n = nilai sampel ke -n
W_n = nilai window ke –n

Jenis *window* yang digunakan adalah *window* Hamming:

$$W_n = 0.54 - 0.46 \cos\left[\frac{2\pi n}{N-1}\right] 0 \ \leq n \leq N-1 \quad (3)$$

4. FFT (*Fast Fourier Transform*)

Untuk mengkonversi setiap *frame* N sampel dari domain waktu ke domain frekuensi. Ketika FFT dilakukan pada *frame*, diasumsikan bahwa sinyal dalam *frame* adalah periodik, dan terus menerus saat membungkus di sekitar. FFT adalah algoritma cepat untuk mengimplementasikan DFT. FFT dihitung dengan persamaan [8]:

$$X_n = \sum_{k=0}^{N-1} X_k e^{-2\pi jkn/N}, \ \ n = 0,1,2,\dots,N-1 \quad (4)$$

Keterangan :
X_n = deretan *aperiodic* dengan nilai *N*
N = jumlah sampel

5. *Mel Filter Bank*

Filter bank adalah teknik filter yang menggunakan representasi konvolusi. Konvolusi dapat dilakukan dengan melakukan multiplikasi antara *spectrum* sinyal dengan koefisien *filter bank*. Formula umum yang digunakan untuk mengkonversi frekuensi ke *mel-scale* dapat pada persamaan:

$$M(f) = 2595 \log 10 \left(1 + \frac{f}{700}\right) \quad (5)$$

Mel filter bank merupakan koleksi filter segitiga yang ditentukan dengan $f_c(m)$, berikut:

$$f(k) < f_{c(m-1)} \quad (5.1)$$

$$H(k,m) = 0 \quad (5.2)$$

$$f_c(m-1) \leq f_{(k)} < f_c(m) \quad (5.3)$$

$$H(k,m) = \frac{f(k) - f_c(m-1)}{f_c(m) - f_c(m-1)} \quad (5.4)$$

$$f_c(m) \leq f_{(k)} < f_c(m+1) \quad (5.5)$$

$$H(k,m) = \frac{f(k) - f_c(m+1)}{f_c(m) - f_c(m+1)} \quad (5.6)$$

$$f(k) \geq f_{c(m+1)} \quad (5.7)$$

$$H(k,m) = 0 \quad (5.8)$$

$$f_c(m) = 700 e^{\left(\frac{\emptyset_C(m)}{1125}\right) - 1} \quad (5.9)$$

Hasil keluaran *Mel Filter bank* diperoleh dengan persamaan (Rinaldi *dkk*, 2016):

$$X'(m) = \ln(\sum_{k=0}^{N-1}|X(k)|H(k,m)) \quad (5.10)$$

6. DCT (*Discrete Cosine Transform*)
   DCT mengimplementasikan fungsi yang sama seperti FFT dengan lebih efisien melalui pengambilan keuntungan dari redudansi yang terdapat dalam sinyal sebenarnya[9].
   $$cn = \sum_{k}^{K}(\log Sk \cos\left[n\left(k-\frac{1}{2}\right)\frac{\pi}{K}\right], n = 1,2,\dots,K) \quad (6)$$

   Sk = keluaran dari proses *filter bank* pada index
   K = jumlah koefisien yang diharapkan

7. *Cepstral Filtering*
   Hasil dari proses utama MFCC memiliki beberapa kelemahan. *Low order* dari *ceptral coefficients* sangat sensitif terhadap *spectral slope* dan bagian *high order* sensitif terhadap *noise. Ceptral filtering* menghaluskan spektrum hasil dari *main processor* sehingga dapat digunakan lebih baik untuk *pattern matching*. Untuk itu dilakukan *cepstral filtering* untuk mengurangi hal-hal tersebut.
   Cepstral filtering dapat dilakukan dengan mengimplementasikan fungsi *window* terhadap *ceptral features* [1].

   $$W_{[n]} = \begin{cases} 1+\frac{L}{2}\sin\left(\frac{nm}{L}\right) & n = 1,2,\dots,L \\ 0 \end{cases} \quad (7)$$

   Keterangan:
   L = jumlah *ceptral coefficients*
   n = *index* dari *ceptral coefficients*

2.1.3 *Dynamic Time Warping* (DTW)

Algoritma DTW didasarkan pada Pemrograman Dinamisteknik digunakan untuk mengukur kesamaan antara dua *time series* yang mungkin bervariasi dalam waktu atau kecepatan. Hal ini kemudian dapat digunakan untuk menemukan daerah yang sesuai antara dua *time series* atau untuk menentukan kesamaan antara dua *time series*. Prinsip dasarnya adalah dengan memberikan sebuah rentang '*steps*' dalam ruang (*frame-frame* waktu dalam sampel, *frame-frame* waktu dalam template) dan digunakan untuk mempertemukan lintasan yang menunjukkan kemiripan antara *time frame* yang lurus. Keunggulan dari DTW adalah dapat menghitung jarak dari dua vektor dengan panjang berbeda. Total *similarity cost* (nilai hasil proses *pattern matching* dua buah suara) yang diperoleh akan menentukan seberapa bagus kesamaan antara *template* dan suara yang diinput.

Jarak antara dua DTW dihitung dari jalur pembengkokkan optimal (*optimal warping path*). Jarak DTW dihitung dengan rumus [1]:

$$D(U,V) = \gamma(m,n) \quad (8)$$

$$\gamma(m,n) = d_{base}(u_i, v_j) + min\begin{cases} \gamma(i-1,j) \\ \gamma(i-1,j-1) \\ \gamma(i,j-l) \end{cases} \quad (9)$$

2.1.4 Normal Distribusi dan MLC (*Maximum Likelihood Classification*)

Normal distribusi berguna untuk distribusi kontinu. Normal distribusi digambarkan seperti kurva berbentuk lonceng di defenisikan dengan *probability density function (pdf)*. Normal distribusi digunakan untuk mewakili variabel-variabel acak [10].

Pada distribusi *Univariate Normal*, hanya menggunakan satu variabel acak yang digunakan sebagai distribusi probabilitas. Sedangkan pada distribusi *Multivariate Normal*, hasil generalisasi dari *univariate* (satu dimensi) ke bentuk dimensi yang lebih tinggi [11].Penentuan parameter distribusi *Multivariate Normal* dilakukan dengan menggunakan metode *Maximum Likelihood* yaitu dengan memaksimumkan fungsi *likelihood* terhadap parameter distribusi. *Maximum Likelihood estimate* akan menemukan parameter terbaik dari parameter yang diberikan [12].

Parameter yang didapat akan digunakan untuk menentukan kelas (*class*). *Maximum Likelihood Classification* adalah teknik untuk evaluasi pelatihan yang digunakan untuk mencari *class* terbaik dari seluruh *class*. Kegiatan evaluasi dilakukan dengan cara $w_i, i = 1, \dots, N$ di mana N adalah jumlah *class*. Kelas yang paling mungkin ditentukan dari vektor *training* data X.

Rumusan umum MLC [11].

$$Class = argmax\ f_i(x, \mu, \theta_i), i \in N \tag{10}$$

Keterangan:
N = jumlah *class*
θ_i = parameter model
x = Data

Transformasi ke dalam bentuk dimensi lain menggunakan model Kernel. Pada model Kernel, parameter yang digunakan adalah parameter yang diberikan oleh data *training* [11]. Bentuk model yang digunakan adalah Polinomial Kernel yaitu $k(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{x}.\boldsymbol{y} + c)^d$ di mana $\boldsymbol{x}$ dan $\boldsymbol{y}$ merupakan vektor baris untuk data. Pengubahan dimensi (d) untuk vektor data dapat dilakukan dengan menggunakan parameter $c = 0$ dan $d = 2$, yaitu dengan mengubah bentuk kernel pada persamaan a ke bentuk parameter tunggal b, yaitu:

$$x = [x_1, x_2] \tag{11}$$
$$y = [y_1, y_2] \tag{11.1}$$
$$k(x, y) = (x_1 y_1 + x_2 y_2) \tag{11.2}$$
$$k(x, y) = (x_1 y_1)^2 + 2x_1 y_1 x_2 y_2 + (x_2 y_2)^2 \tag{11.3}$$
$$k(x, y) = \begin{bmatrix} {x_1}^2 \\ \sqrt{2x_1 y_1} \\ {x_2}^2 \end{bmatrix} \begin{bmatrix} {y_1}^2 \\ \sqrt{2x_2 y_2} \\ {y_2}^2 \end{bmatrix} \tag{11.4}$$

sehingga didapat:

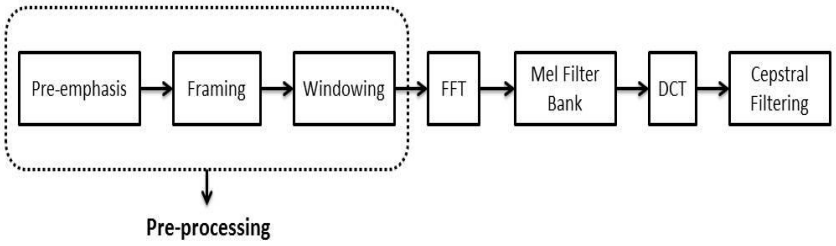
$$\phi(\boldsymbol{x}) = \begin{bmatrix} {\mu_1}^2 \\ \sqrt{2\pi_1 \mu_2} \\ {\mu_2}^2 \end{bmatrix} \tag{11.5}$$

2. 2 Pengumpulan Data

Data rekaman diambil dari 20 penutur (10 laki-laki dan 10 perempuan). Setiap penutur akan mengucapkan kalimat bebas dalam Bahasa Indonesia yang diambil dari buku, novel, dan majalah. Panjangnya kalimat tergantung dari kecepatan setiap penutur berbicara. Penutur akan mengucapkan sebanyak 6 kali untuk data training, masing-masing 30 detik. Rekaman akan dilakukan dengan 3 kondisi perpustakaan yaitu sepi, noiseless, dan noise dengan musik. Setiap kondisi ruangan akan dilakukan rekaman sebanyak 2 kali. Suara akan direkam menggunakan fitur yang disediakan oleh MATLAB. File suara akan tersimpan dalam format .wav dengan ukuran 16 *bit* dan *rate* 8 KHz.

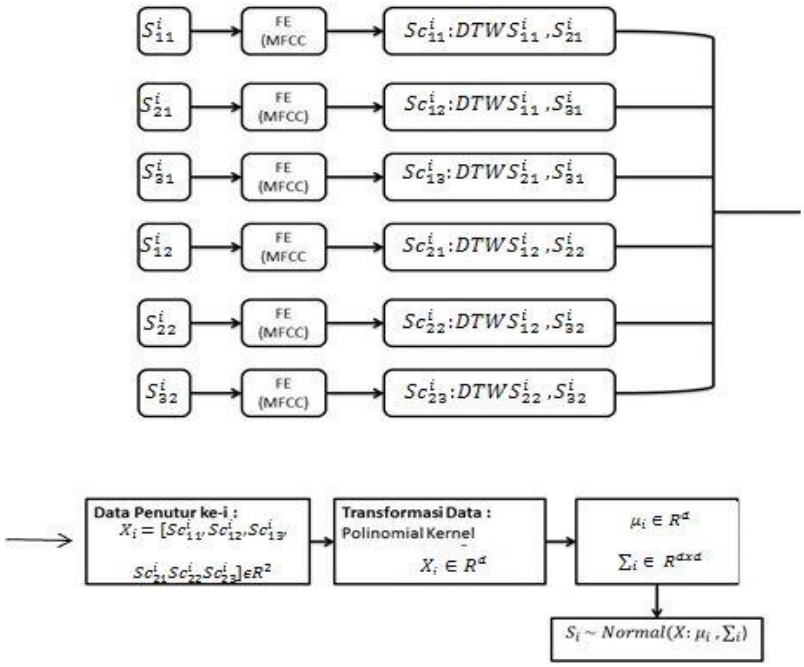
2. 3 Perancangan

Rancangan sistem pada penelitian ini terdiri dari proses *input* suara, *pre-processing*, *feature extraction* (FE) suara dengan MFCC, dan pengklasifikasian suara dengan algoritma DTW. Setiap suara akan diekstrak menggunakan MFCC dan diolah dengan suara lain untuk mengambil ciri dari suara. Tahapan fitur ekstrasi MFCC dapat dilihat pada Gambar 1.



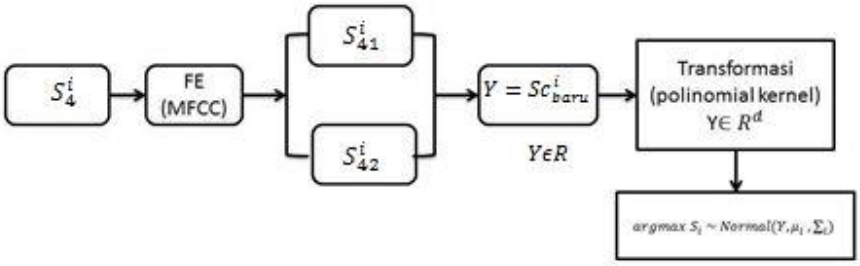
Gambar 1. Tahap MFCC

Setelah didapatkan 6 buah *similarity score* dari berbagai kombinasi data sampel, nilai-nilai tersebut ditransformasi dengan metode Polinomial Kernel $(\boldsymbol{x}.\boldsymbol{y})^2$. Berikut gambar tahapan pada saat *training*.



Gambar 2. Tahap *Training*

Selanjutnya akan dilakukan proses *testing*. Data diambil dari 20 penutur masing-masing mengucapkan kalimat bebas dengan durasi 60 detik. Suara tersebut akan dibagi menjadi 2, masing-masing dengan durasi 30 detik. Sama seperti proses *training*, data suara akan melalui *pre-processing*, *feature extraction* dan perhitungan *similiarity score*. Hasil *similiarity score* tersebut akan dibandingkan dengan rata-rata skor dan standar deviasi dari setiap *similiarity score* penutur.



Gambar 3. Tahap *Testing*

3. HASIL DAN PEMBAHASAN

Gambar 4 menunjukkan tampilan proses untuk merekam suara dan proses ekstraksi ciri sampai pengklasifikasian suara. Nama file yang disimpan setelah merekam akan tertera pada tampilan. Selanjutnya akan dilakukan proses ekstraksi ciri dan klasifikasi dengan menekan tombol "Proses". Hasil akan tertera pada kotak kecil. Suara yang akan diproses juga dapat langsung dipilih dari *file* yang telah tersedia.

Gambar 4. Tampilan Rekam dan Pengklasifikasian Suara

3. 1 Analisis Hasil

Sebelum dilakukan pengujian, terlebih dahulu dilakukan *training* pada data set yang sudah dikumpulkan. Proses *training* digunakan untuk melatih mesin agar dapat melakukan pengenalan suara yang baru ditambahkan. Selanjutnya dilakukan pengujian untuk mendapatkan data *testing***.** Hasil data *training* dan *testing* akan ditransformasikan

ke bentuk dua dimensi menggunakan polinomial kernel untuk mendapatkan *probability* lain yang dapat dijadikan sebagai ciri (data) baru.

Tahap *training* dan *testing* diawali dengan mencari ciri suara menggunakan *feature extraction* MFCC. MFCC menghasilkan matriks 13x2998 di mana 2998 merepresentasikan koefisien yang dihasilkan per-*frame* suara dan 13 merupakan gabungan koefisien semua *frame* dari keseluruhan sampel rekaman. Pada tahap training akan menghasilkan 6 *similiarity score* yang didapatkan dari 6 data suara setiap penutur. Dari data *training* akan didapatkan parameter model berupa rata-rata dan standar deviasi. Model tersebut akan digunakan untuk membandingkan dengan hasil dari tahap *testing*.

Pengujian performa dari MFCC dan algoritma DTW akan dilakukan untuk identifikasi dan verifikasi suara. Pada identifikasi akan dicari nilai maksimal dari seluruh *score* menggunakan *Maximum Likelihood Classifier* (MLC). Pada verifikasi menggunakan konsep Operasi Ambang Batas (*Thresholding*) dengan menjadikan *score* dari penutur yang akan diuji sebagai syarat ambang batas.

Pengujian data *training* (*recall*) untuk verifikasi dilakukan dengan membandingkan setiap 6 skor milik suara ke-n dengan 6 skor yang telah melewati normal distribusi dari 19 suara lainnya. Nilai yang memenuhi syarat ambang batas akan dipetakan ke satu nilai yang dikehendaki. Jika skor suara lain memenuhi syarat *threshold* terhadap skor suara ke-n yang diuji, maka setiap skor akan dipetakan ke skor suara ke-n. Setiap skor suara yang memenuhi syarat akan dijumlahkan sesuai dengan berapa kali skor suara tersebut berhasil memenuhi syarat nilai ambang atas saat dibandingkan pada 19 suara lainnya. Jumlah maksimal dari keseluruhan adalah 19.

Pada pengujian verifikasi untuk satu dimensi didapatkan akurasi sebesar 77,3% dengan 7 penutur mendapatkan presentase 100%, sedangkan untuk dua dimensi didapatkan akurasi sebesar 84,9% dengan 13 penutur yang mendapat persentase sebesar 100%.

Pengujian *recall* untuk identifikasi dilakukan dengan mengambil nilai maksimal 6 data *training* terhadap 20 suara. Pada pengujian identifikasi untuk satu dimensi didapatkan akurasi sebesar 15,8% dan untuk dua dimensi didapatkan akurasi sebesar 17,4%.

Pengujian data *testing* (*precision*) menggunakan data baru yang belum pernah digunakan pada tahap *training*. Pengujian ini memerlukan model berupa rata-rata dan standar deviasi per suara. Pada verifikasi akan ditentukan dengan membandingkan 20 model suara *training* dan pada identifikasi ditentukan dengan mengambil skor terbesar pada setiap suara yang dites.

Pada pengujian *precision* verifikasi untuk satu dimensi dan dua dimensi didapatkan akurasi sebesar 70% dengan 14 penutur yang berhasil diverifikasi. Pada pengujian *precision* identifikasi untuk satu dimensi dan dua dimensi didapatkan akurasi sebesar 15% dengan 3 penutur yang berhasil diidentifikasi.

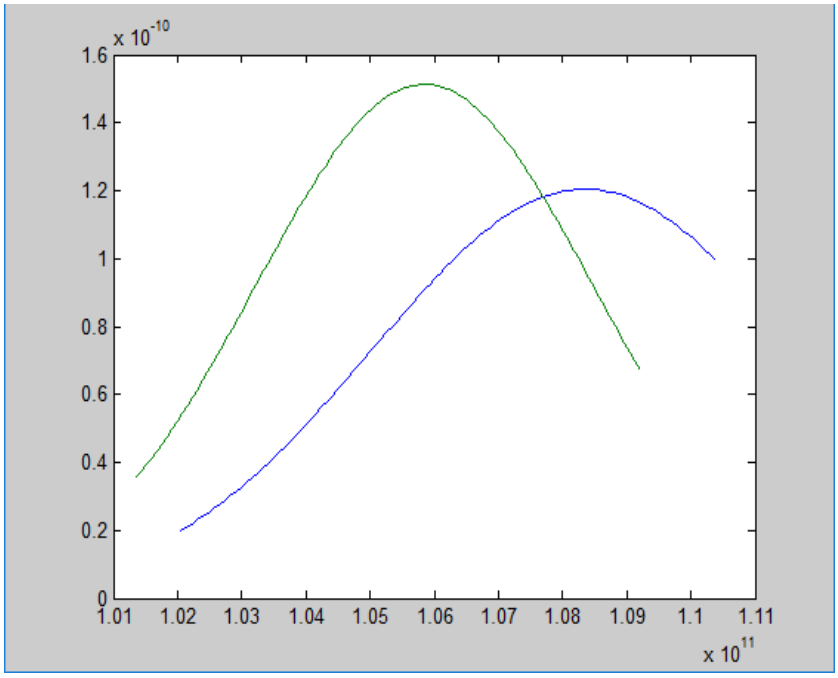
Tabel 1. Hasil Pengujian

Dimensi	**Persentase Verifikasi**		**Persentase Identifikasi**	
	Recall	***Precision***	***Recall***	***Precision***
Satu	77,73%	70%	15,8%	15%
Dua	84,9%	70%	17,4%	15%

Berdasarkan hasil pada Tabel 1, keberhasilan pengujian *recall* dan *precision* pada data dua dimensi lebih besar daripada data satu dimensi. Hal tersebut menunjukan bahwa ciri (data) terbaik didapatkan setelah dilakukan tranformasi data ke dimensi yang lebih tinggi sehingga menambah *probability* yang lain.

Pada pengujian untuk verifikasi dan identifikasi, didapatkan hasil yang berbeda. Persentase keberhasilan pada verifikasi lebih besar dan lebih baik dibandingkan identifikasi. Hal ini menunjukkan bahwa sistem mampu mengecek kebenaran data yang baru dimasukkan terhadap data yang sudah ada.

Hasil persentase identifikasi lebih kecil dikarenakan saat dilakukan tes, sistem akan mengambil nilai tertinggi dari seluruh sampel suara penutur. Sedangkan ada penutur yang keseluruhan skornya lebih tinggi dari pada keseluruhan skor penutur-penutur lain sehingga suara yang akan diidentifikasi sering kali salah. Salah satu contohnya adalah pada suara ke-1 dan ke-11 sebagai berikut.



Gambar 5. Grafik Suara ke-1 dan Suara Ke-11

Grafik pada Gambar 5 menunjukkan suara ke-1 (biru) dan suara ke-11 (hijau) dalam bentuk satu dimensi di mana jika terdapat suara dengan nilai $x = 1.06\times10^{11}$ maka label yang dipilih adalah untuk suara 11 karena memiliki nilai probabilitas yang tinggi (*local maxima*). Hal ini bermasalah ketika data tersebut adalah suara ke-1, yaitu kegagalan menemukan karakteristik data.

4. KESIMPULAN

Dari hasil penelitian dan pembahasan yang dilakukan maka dapat disimpulkan beberapa hal sebagai berikut:

1. MFCC dapat digunakan dalam pengekstraksian ciri suara.
2. MFCC dan DTW memiliki nilai akurasi yang kecil pada identifikasi suara yaitu 17,4% untuk pengujian *recall* dan 15% untuk pengujian *precision*, sedangkan untuk verifikasi memiliki hasil akurasi yang lebih baik yaitu 84,9% untuk pengujian *recall* dan 70% untuk pengujian *precision*.
3. Transformasi data menjadi dua dimensi dapat meningkatkan persentase keberhasilan.

5. SARAN

Saran yang diberikan untuk penelitian selanjutnya adalah:

1. Peningkatan akurasi dari MFCC dan DTW untuk verifikasi maupun identifikasi suara dapat ditingkatkan dengan menggunakan kernel lain untuk transformasi data, seperti Gaussian Kernel, Sigmoid Kernel, Invers Multi Kuadratik, dll.
2. Penggunaan metode reduksi data dapat mengurangi waktu kerja MFCC dan DTW.

DAFTAR PUSTAKA

[1] Putra, D & Adi R 2011, Verifikasi Biometrika Suara Menggunakan Metode MFCC dan DTW, *Lontar Komputer*, Vol. 2, No. 1, h. 8-21, Denpasar.

[2] Mohammad, FR.,dkk 2011, A Hybrid Reliable Algorithm for Speaker Recognition based on Improved DTW and VQ by Genetic Algorithm in Noisy Environment, *International Conference on Multimedia and Signal Processing*, Vol. 2, h. 269-273, Iran.

[3] Nidyananthan, SS & R. Shantha Selva Kumari 2013, Language and Text Independent Speaker Identification Systems using GMM, *WSEAS Transactions on Signal Processing*, Vol. 9, No. 4, h. 185-194 , India.

[4] Syah, DPA 2009, Sistem Biometriks Absensi Karyawan Dalam Menunjang Efektifitas Kinerja Perusahaan, http://donupermana.wordpress.com/ makalah/sistem-biometrik-absensi/, diakses tgl 08 September 2016.

[5] Sharma, P 2013, Adult Voice Recognition System using Text Variable Phoneme Model and Coarse Speaking Fundamental Frequency Characteristics, *The International Journal of Engineering and Science (IJES)*, Vol. 2, No.1, h. 126-132, Rajpura.

[6] Paul, S.,dkk 2015, Text Dependent Speaker Verification using Algebraic Approach (AA) method and DTW under limited data condition, dari ieeexplore.ieee.org, India.

[7] Muda, L.,dkk 2010, Voice Recognition Algorithms Using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Technique, *Journal of Computing*, Vol. 2, No. 3, h 138-143, Malaysia.

[8] Mustofa, A 2008, Sistem Pengenalan Penutur dengan Metode Mel-Frequency Wraping, *Jurnal Teknik Elektro*, Malang.

[9] Rinaldi, A.,dkk 2014, Pengenalan Gender Melalui Suara dengan Algoritma Support Vector Machine (SVM), dari eprints.MDP.ac.id, Palembang.

[10] Keith Dean Simonton, "*Distribution, Normal"*, http://www.encyclopedia.com/science-and-technology/mathematics/mathematics/normal-distribution#3, diakses tgl 21 November 2016.

[11] Sugiyama, M 2016, *Introduction to Statistical Machine Learning*, Elsevier, h. 157-159, USA.

[12] Budi, IS 2014, Membangun *Gaussian Classifier* Dalam Mengenal Objek Dalam Bentuk *Image*, dari ejournal.uin-malang.ac.id, Vol.1, No.1, h.21-26, Malang.